

ISTA 421 / INFO 521 – Final Project OPTION A: Metropolis-Hastings MCMC Inference of 3D Line

Due: Friday, December 8, 8pm

Total 20% of final grade

STUDENT NAME

Undergraduate / Graduate

Introduction

The following are the instructions for Option A of the final project. If you are doing Option B, you do NOT do this option; you should instead refer to the instructions for Option B.

The instructions start with a general background description of the model and problem scenario, followed by a set of Tasks. You must commit and push your executable code along with a PDF of written answers to your final_project repository. The name of your written PDF must be `final_project.OptionA.answers.pdf`. The first task has you implement the core code to your Metropolis-Hastings sampler. Tasks 2 through 5 then require you to run that code on the provided data in various configurations, generate plots, report values, and describe the outcomes.

Option A Background: Inferring parameters of a 3D line from noisy 2D images using Metropolis-Hastings MCMC

In vision problems, we often represent the process by which an image is produced by considering a parameterized camera model (in a sense, the camera model defines the “perspective” toward the scene) and the scene being imaged.

In the pinhole camera model, cameras can be represented by matrices that contain parameters for the camera, such as its position in space and its focal length.¹ This representation is useful because it allows us to obtain 2D images of 3D points in a very simple way.

Let $\mathbf{M} \in \mathbb{R}^{3 \times 4}$ be a camera matrix. Then, if $\mathbf{p} \in \mathbb{R}^3$ is a 3D point (i.e., $\mathbf{p} = [p_1, p_2, p_3]^\top, p_i \in \mathbb{R}$), its 2D image $\mathbf{q} \in \mathbb{R}^2$ is given by

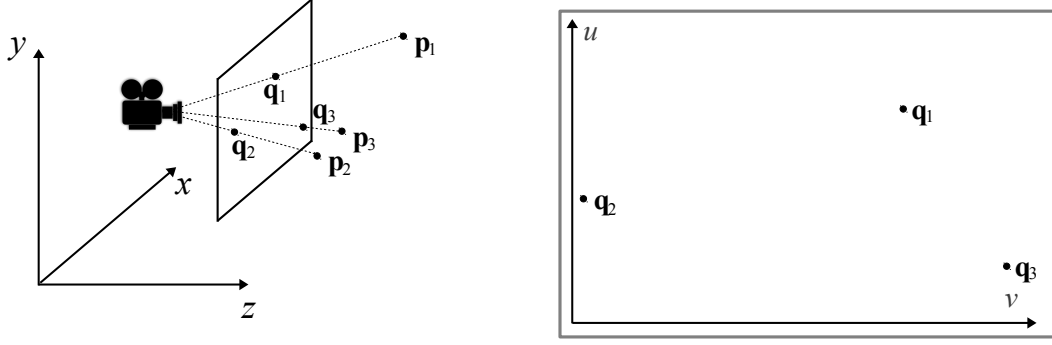
$$\mathbf{q} = \begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{\tilde{w}} \begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} \quad (1)$$

where \tilde{u} , \tilde{v} , and \tilde{w} are defined by

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \mathbf{M} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} = \mathbf{M} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ 1 \end{bmatrix} \quad (2)$$

Notice the number 1 that is appended to \mathbf{p} when multiplying by \mathbf{M} . This is called the *homogeneous* coordinate, and it is necessary for posing camera projections as matrix multiplication. This projection principle is illustrated in Figure 1

¹For those interested (*not* required for this project), details can be found in Hartley and Zisserman (2004) *Multiple View Geometry in Computer Vision*; you can also internet search for “camera matrix.”



(a) A depiction of the projection of three points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \in \mathbb{R}^3$ onto the image plane of a camera using Eqs 1 and 2.

(b) Graphical depiction of the 2D image plane with image points $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3 \in \mathbb{R}^2$ projected from points $\mathbf{p}_1, \mathbf{p}_2$, and \mathbf{p}_3 .

Figure 1: Two views of projection.

NOTE: A complete understanding of this material, while very valuable in itself (!), is not necessary to complete this problem. The only thing you are required to understand is that 3D points (in coordinates x, y and z of Figure 1a) become 2D image pixel points (in coordinates u, v of Figure 1b) by Equations 1 and 2.

We will start by defining a “Camera 1” whose matrix is

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Camera 1 corresponds to a camera located at world coordinates $(0,0,0)$ and pointing down the positive z -axis, with unit focal length. What follows is the problem setup.

Let $\ell = (\mathbf{p}_i, \mathbf{p}_f)$ be a 3D line segment with end-points $\mathbf{p}_i, \mathbf{p}_f \in \mathbb{R}^3$. (The subscripts on the end-points stand for “initial” and “final”, respectively.) Now, consider the following generative process. We “take a picture” of ℓ using camera \mathbf{M} , producing $\ell_{2D} = (\mathbf{q}_i, \mathbf{q}_f)$ (that is, the \mathbf{q}_i and \mathbf{q}_f are obtained using Equations 1 and 2).

We then sample S points \mathbf{q}_s from the line ℓ_{2D} . This is accomplished by taking the ratios t_1, \dots, t_S along the straight-line path from \mathbf{q}_i to \mathbf{q}_f so that $\mathbf{q}_s = \mathbf{q}_i + (\mathbf{q}_f - \mathbf{q}_i)t_s$, with $0 \leq t_s \leq 1$, $s = 1, \dots, S$. (Figure 2(b) shows one of these sampled points, \mathbf{q}_m , which happens to be exactly half way between \mathbf{q}_i and \mathbf{q}_f , so for this point $t_m = 0.5$).

In our scenario, the imaging and rendering processes are inherently noisy, so we add Gaussian noise to the \mathbf{q}_s , which gives us $\mathbf{r}_s = \mathbf{q}_s + \epsilon$, where $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma)$, with Σ a 2×2 covariance matrix (since this noise is in the 2D image plane). Naturally, this implies that $\mathbf{r}_s | \mathbf{p}_i, \mathbf{p}_f, t_s, \Sigma \sim \mathcal{N}(\mathbf{q}_s, \Sigma)$. See Figure 2 for a summary illustration of the generative process.

We will additionally place a Gaussian prior on the possible end-points, $\mathbf{p}_i, \mathbf{p}_f$, of the line ℓ (this is a *different* Gaussian than the image Gaussian noise ϵ). These prior parameters are subscripted by the bold- ℓ to indicate they refer to the properties of the 3D line. This represents the prior belief about where the line, as defined by end-points \mathbf{p}_i and \mathbf{p}_f , might be in 3D space:

$$\mathbf{p}_i, \mathbf{p}_f \sim \mathcal{N}(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell).$$

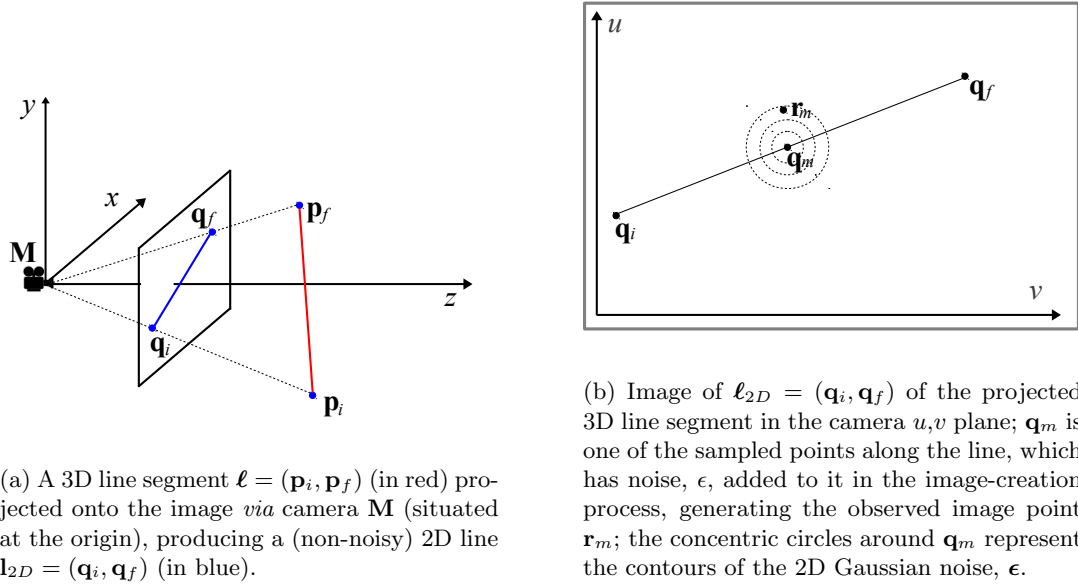


Figure 2: The noisy generative process projecting end-points along the line to the image.

Tasks [34 points total]

All data files needed for this task are provided in `projectA-3D-line-MH/data_2023/`

For this project option, you will need to complete the following five tasks:

1. [10 points] Write a python script that, given
 - inputs t_1, \dots, t_S (the relative offset ratios of points \mathbf{q}_s along the line segment between the \mathbf{q}_i and \mathbf{q}_f endpoints),
 - 2D points $\mathbf{r}_1, \dots, \mathbf{r}_S$ (the observed points in the camera u, v image plane),
 - Σ (the image projection noise covariance),
 - μ_ℓ and Σ_ℓ (the priors over the 3D line segment endpoints),

uses the Metropolis-Hastings (M-H) algorithm to sample from the posterior of \mathbf{p}_i and \mathbf{p}_f . In other words, the script should generate samples of \mathbf{p}_i and \mathbf{p}_f according to the posterior distribution

$$p(\mathbf{p}_i, \mathbf{p}_f | \mathbf{r}, \mathbf{t}, \Sigma, \mu_\ell, \Sigma_\ell) \propto p(\mathbf{r} | \mathbf{p}_i, \mathbf{p}_f, \mathbf{t}, \Sigma) p(\mathbf{p}_i, \mathbf{p}_f | \mu_\ell, \Sigma_\ell),$$

with $\mathbf{r} = \{\mathbf{r}_1, \dots, \mathbf{r}_S\}$ and $\mathbf{t} = \{t_1, \dots, t_S\}$.

NOTE: The prior and likelihood probabilities can be *extremely* small, to the point that their product can result in floating-point number underflow errors – i.e., the computer cannot represent the small values and so treats them as though they were zero. This particularly becomes a hazard when you take the product of two very small numbers, which can happen when computing the posteriors in the M-H acceptance ratio, which in turn can lead to “division by zero” errors (when the denominator of the acceptance ratio is treated as zero). A work-around is to take the log (remember, we follow the machine learning convention that this is really the *natural log*) of the acceptance ratio (be sure to compute the log completely through the ratio and products), do your numeric computations to get the log-ratio α . You then have two options for how you can use this log-ratio α :

- You can transform the final value back into the original scale by computing e^α , and then compute your acceptance test;
- *or* keep it in the log scale. If you do this, you must also ensure your random sample is also log scale! For this, the python function `scipy.stats.multivariate_normal.logpdf` will be handy. Also, keep in mind that when you perform your acceptance test, that must also be considered in the log-scale.

The written answer for this task (task 1) requires that you describe how to run your script for the following 4 tasks (tasks 2 through 5). As you complete the following tasks, you should update your description here.

2. **[6 points]** Use your script to draw samples from the posterior distribution of the 3D line segment end-points.

The observations you will use in this task are taken from the perspective of Camera 1 only. As described above, the observations consist of points \mathbf{r}_s in the u, v image plane of the camera that result from a noisy projection process originating from points \mathbf{q}_s . Each \mathbf{q}_s was itself sampled some relative offset t_s (relative to the segment endpoints, \mathbf{q}_i and \mathbf{q}_f). The data file `inputs.csv` contains the relative offset ratios for each \mathbf{q}_s . The projection of point \mathbf{q}_s , however, is not directly observed, but instead appears in the u, v image plane as the result of a noisy process (Figure 2(b)) that produces the observed point \mathbf{r}_s in the u, v image plane coordinates. The u, v coordinates of each of these observations are contained in the data file `points_2d_camera_1.csv`. (The row indices of the offset ratios in `inputs.csv` corresponds to the row indices of the points in `points_2d_camera_1.csv`.)

Assume the following prior parameters: $\Sigma_\ell = 6 \cdot \mathbf{I}_{3 \times 3}$ (where here $\mathbf{I}_{3 \times 3}$ is the 3×3 identity matrix), $\mu_\ell = [0, 0, 6]^\top$.

Also assume that $\Sigma = (0.05)^2 \cdot \mathbf{I}_{2 \times 2}$ (where here $\mathbf{I}_{2 \times 2}$ is the 2×2 identity matrix and $(0.05)^2$ is the scalar value of the square of 0.05).

It is recommended that you sample your initial end-points from the prior. Sample for 50,000 iterations.

It is up to you to choose your proposal distribution! You may need to try some different parameterizations to get a good one.

Note: you will likely find your best (MAP) fit points within the first 3,000-5,000 samples, but you should sample for 50,000 to see what the longer-term trend in your estimates looks like.

In your written answers for this task, including the following:

- (3 points) Generate a plot like that of Figure 4.12(c), page 161 of FCML (2nd edition), for each parameter you are estimating, to view progress during sampling. Remember, only plot the *accepted* samples, not the rejected samples. Provide informative captions.
 - (3 points) Describe the trends you see in these plots, and include an explanation of what they are telling you. Expect to see some differences between the plots you generate and what you see in FCML Figure 4.12(c).
3. **[4 points]** Using the samples you generated, find the maximum a posteriori (MAP) estimate of the line end-points. Report your results, including:
 - (1 point) Report your MAP estimate of the line end-points (the 3d points of each endpoint).
 - (3 points) Plot the MAP estimate of the line (that is, the MAP for each endpoint with a line connecting them) as projected onto the Camera 1 image plane (using the Camera 1 camera matrix). Include in the plot the original noisy observations, \mathbf{r} , of points sampled along the true line. You should see that your inferred line fits fairly well among the points. If not, try running your M-H procedure again with a different starting point for your line end-points (it is generally a

good idea to sample your first end-points from the prior; update your plots in task 2 if you re-run, so that what you report there corresponds to the same data that you report here). Include in your caption an explanation for what is plotted.

4. [4 points] Suppose we have a second camera

$$\mathbf{M}' = \begin{bmatrix} 0 & 0 & 1 & -5 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 5 \end{bmatrix},$$

(a camera at $[5, 0, 5]^\top$ looking in the negative x -axis and unit focal length), and, consequently, a second set of observed data points $\mathbf{r}'_1, \dots, \mathbf{r}'_S$, where the indices of these points correspond to the same points viewed from the first camera (those were provided in `points_2d_camera_1.csv`), but now we have what those points look like through the second camera, as found in `points_2d_camera_2.csv`. First, let's see how your estimate from task 2 looks from this second camera:

- Create another plot like you did in task 3, where you plotted the MAP hypothesis of the end-points sampled in task 2, but instead use the second camera matrix to project your MAP hypothesis of the end-points (still connect them by a line), and also plot the new views of the points along the line from the perspective of the second camera, as found in `points_2d_camera_2.csv`. Include in your caption an explanation of what is plotted, and how it compares to the plot for Camera 1.
5. [10 points] Now, perform M-H sampling again, but this time use the data from **both** cameras. You will need to adapt your likelihood to now consider the two camera data sources, but the rest of M-H stays the same! Again, 50,000 iterations in M-H should be sufficient to converge and to see the longer-term trend. Report the following:
- (2 points) Find and report the MAP estimate of the line using the same setup as in task 2 using both the observations in `points_2d_camera_1.csv` from the first camera and the points `points_2d_camera_2.csv` viewed by the second camera.
 - (4 points) Make the same plots that you did in task 2. Do they qualitatively look the same as those from task 2? If they're different, describe how and offer an explanation as to why we see those differences.
 - (4 points) Plot the projection of your **new** MAP end-points estimate for the combined cameras; make one plot each for the projection of your MAP end-points connected by a line for Camera 1 (along with the noisy data for that camera), and another for Camera 2. How well do you fit the points in both camera projections?